

(9) are marked prosodically, but this is a matter of stylized tunes which affects only isolated utterances. Note that in example (4) it is the fact that the professor and the student did sustain a cooperative exchange after the opening sequence that led the student to misread the professor's intent. Similarly, a native American who differed from the black speaker in (8), and consequently failed to understand that a simile was intended, might nevertheless have realized that rising intonation indicated the speaker was not ready to relinquish her turn.

There is reason to believe that the differences between Western – i.e. native British and American – and Indian English are matters of basic cultural norms and of the interaction of prosody and syntax reflecting long established, historical traditions that arose in distinct culture areas, and are maintained through networks of interpersonal relationships. Individuals reared in these traditions often learn the clause level grammar of another language, but in using it they rely on their own native discourse conventions. These conventions, as was argued in chapters 4 and 5, are subconscious and for the most part tend to remain unverbalized. They are learned only through prolonged and intensive face to face contact. Yet the very linguistic features that cause the comprehension problem also make it difficult to enter into the type of contact and elicit the type of feedback that is necessary to overcome them. In this way casual intergroup contacts may reinforce distance and maintain separateness unless stronger outside forces intervene to create the conditions that make intensive interaction possible.

Discourse strategies

7 Gumperz, John 1982

Socio-cultural knowledge in conversational inference

IN DISCOURSE STRATEGIES

JOHN GUMPERZ 1982

Conversational inference, as I use the term, is the situated or context-bound process of interpretation, by means of which participants in an exchange assess others' intentions, and on which they base their responses.

Recent studies of conversation from a variety of linguistic, psychological, anthropological and sociological perspectives, have shed light upon a number of issues important to the study of conversational inference. It is generally agreed that grammatical and lexical knowledge are only two of several factors in the interpretation process. Aside from physical setting, participants' personal background knowledge and their attitudes toward each other, socio-cultural assumptions concerning role and status relationships as well as social values associated with various message components also play an important role. So far, however, treatment of such contextual factors has been primarily descriptive. The procedure has been to identify or list what can potentially affect interpretation. With rare exceptions, no systematic attempts are made to show how social knowledge is used in situated interpretation. Yet we know that social presuppositions and attitudes shift in the course of interaction, often without a corresponding change in extralinguistic context. As we have argued in previous chapters, the social input to conversation is itself communicated through a system of verbal and nonverbal signs that both channel the progress of an encounter and affect the interpretation of intent. It follows that analysis of such ongoing processes requires different and perhaps more indirect methods of study which examine not the lexical meanings of words or the semantic structure of sentences but interpretation as a function of the dynamic pattern of moves and countermoves as they follow one another in ongoing conversation.

Conversational inference is part of the very act of conversing. One indirectly or implicitly indicates how an utterance is to be interpreted and illustrates how one has interpreted another's utterance through verbal and nonverbal responses, and it is the nature of these responses rather than the independently determined meaning or truth value of individual utterances alone that governs evaluation of intent. This chapter suggests the outlines of a theory that deals with the question of how social knowledge is stored in the mind, how it is retrieved from memory and how it interacts with grammatical and lexical knowledge in the act of conversing. To put the discussion in context, we will begin with a brief outline of some of the major research traditions that deal with contextual factors in interpretation. We will then go on to analyze several brief conversational exchanges illustrating various aspects of the inferential process.

Ethnography of communication and discourse analysis

Existing theories visualize the relationship of extralinguistic, socio-cultural knowledge to grammar in one of two ways. The first is the anthropological tradition of ethnography of communication, where socio-cultural knowledge is seen as revealed in the performance of speech events defined as sequences of acts bounded in real time and space, and characterized by culturally specific values and norms that constrain both the form and the content of what is said. The second tradition of discourse analysis, deriving from speech act theory, linguistic pragmatics, frame semantics (Fillmore 1977) and artificial intelligence posits abstract semantic constructs, variously called scripts, schemata, or frames, by means of which participants apply their knowledge of the world to the interpretation of what goes on in an encounter. The two traditions differ both in theory and in methodological approach.

Although ultimately concerned with communicative competence, i.e. abstract cognitive knowledge, the initial goal of ethnography of communication is, as Hymes (1962) puts it, "to fill the gap between what is usually put into ethnography and what is usually put into grammar."

It is argued that because of the linguist's concern with historical reconstruction and context free grammatical rules, existing grammars are built on a highly selective data base and do not provide the information needed for understanding how language is employed.

New types of data are needed. Theoretical writings in the ethnography of speaking seek to fill this need and are in large part programmatic, suggesting categories of inquiry intended to guide empirical data selection. Studies of language use are called for which concentrate on what Hymes calls the *means of speaking*. This includes information on the local *linguistic repertoire*, the totality of distinct language varieties, dialects and styles employed in a community. Also to be described are the *genres* or art forms in terms of which verbal performances can be characterized, such as myths, epics, tales, narratives and the like. Descriptions further cover the various acts of speaking prevalent in a particular group ('act' is used here broadly, in Austin's sense, to suggest functions such as question, response, request), and finally the 'frames' that serve as instructions on how to interpret a sequence of acts (Bauman & Sherzer 1975).

The means of speaking are put into practice and related to cultural norms in the performance of particular speech events. Action in such events is seen as governed by social norms specifying such things as who can take part, what the role relationships are, what kind of content is admissible, in what order information is to be introduced, and what speech etiquette applies. To describe these norms, the ethnographer relies on the usual anthropological field methods.

Ethnographers of communication have collected new, highly valuable descriptive information documenting the enormous range of signalling resources available in various cultures, as well as many culturally specific ways that rules of speaking vary with context. They have provided convincing evidence to show that much of language use, like a grammar, is rule governed. In specifying what these rules are, they have rejected the traditional functionalist paradigms in which languages and cultures are seen as separate unitary wholes, but they tend to see speech events as bounded units, functioning somewhat like miniature social systems where norms and values constitute independent variables, separate from language proper. The task of sociolinguistic analysis, in this view, is to specify the interrelationship of such variables in events characteristic of particular social groups. The question of how group boundaries can be determined, is not dealt with, nor are the issues of how members themselves identify events, how social input varies in the course of an interaction and how social knowledge affects the interpretation of messages. The principal goal is to show how social norms affect the

use and distribution of communicative resources, not to deal with interpretation.

In the second of our two traditions, that of discourse analysis, the cognitive functioning of contextual and other knowledge becomes the primary concern. Initially, work in this tradition was motivated in large part by a concern with basic grammatical and semantic theory. In a sense it can be seen as an effort to give linguistic substance to Wittgenstein's and Austin's philosophical writings, which point to the inadequacies of the logician's concept of meaning as the relationship of words or sentences to things or ideas and argue that meaning ultimately resides in human action. The key notion is Grice's (1957) definition of meaning as "the effect that a sender intends to produce on a receiver by means of a message." Speech acts defined in terms of illocutionary force, i.e. utterers' communicative intent, become the main unit of linguistic analysis (Grice 1957, 1971).

As in Chomskian generative grammar, analysis focuses on what speakers must know in order to identify such acts as, for example, declaratives, questions, requests, or suggestions. It is agreed that speech act interpretation always relies on extralinguistic presuppositions, along with grammatical knowledge. In attempting to specify what these presuppositions are, research has increasingly come to concentrate on text comprehension rather than on sentences as such. The view here is, however, basically a psycholinguistic one of individual members of a culture speaking a specific language or dialect, drawing on their *knowledge of the world* to interpret utterances in context. Various mechanisms have been proposed for describing the cognitive structures involved and showing how they can enter into interpretation. Cognitive psychologists and specialists in artificial intelligence tend to work deductively, starting out with formalizable constructs like schemata, scripts and plans (Bobrow & Collins 1975, Schank & Abelson 1977) that reflect knowledge relevant to common discourse situations like eating in a restaurant or getting information about plane travel. These constructs are seen to function somewhat like the plot of a play, which specifies goals and subunits of an action, as well as relationships among acts, and provides and enables the audience to fill in outside information not specified in the overt content of messages.

A related view of world knowledge is reflected in Fillmore's (1977) concept of 'scene,' where meaning is characterized iconically rather

than in terms of lexical sequences or abstract semantic formalisms. Scenes are like pictures, in that they can be described from various perspectives and from differing participants' points of view. Relevant aspects of meaning are signalled partly through lexical meaning and partly through syntactic or prosodic channels. Presumably once readers or listeners have read or heard enough to form hypotheses of what schemata are involved, these hypotheses then supply the world knowledge needed to fill in nonverbalized information. This iconic view of interpretation is particularly important from a sociolinguistic perspective, since it can be shown that the signalling load which the particular linguistic channels carry in depicting scenes varies from language to language, so that referentially similar messages can be interpreted differently by individuals who approach the message with differing presuppositions.

Although the two research traditions differ both in theory and in methodological approach, they share similar notions as to what linguistic signalling mechanisms are. Both define the basic theoretical issue as one of showing how extralinguistic knowledge, reflected in cognitive or social structures that exist independently apart from communication, are brought into the speech situation. Where discourse is analyzed, the aim is to produce ideotypical descriptions that can be dissected into significant components and used to produce typologies. It is these typified, generalized structures that are then used to explain what happens in everyday situations.

Structural analyses of events or interpretive schemata have furnished proof that interpretation is context bound and that human knowledge is best treated as situation specific. Yet any attempt to apply such ideotypical constructs in the study of everyday verbal exchanges is certain to encounter serious problems. To begin with, although event labels and discourse categories are part of our everyday vocabulary and are regularly used when we talk *about* modes of speaking, they are highly abstract in nature and on the whole poor descriptors of what is actually accomplished. When participants report on actual verbal encounters, they tend to do so by mentioning some item of content, or by referring to what people were getting at or what they were trying to do. Event names in everyday talk are most often used metaphorically to refer retrospectively to what was accomplished.

If I say to someone "I think we need to have a chat," the activity I intend to engage in is quite unlikely to be chatting. Nor is it always possible to predict what is intended simply by specifying what we as members of the culture know about the extralinguistic setting, personal desires of participants and the content of what has transpired. The discussion of interpretive issues in previous chapters indicates that situated interpretations are problematic and not equally available to those who know the context and can decode isolated sentences, so we need to examine interaction itself to learn how contextual presuppositions function.

Conversational analysts concerned with naturally occurring instances of everyday talk follow still another, separate academic tradition of inquiry, which concentrates on the actual discourse mechanisms that serve to allocate turns of speaking, to negotiate changes in focus and to manage and direct the flow of interaction, and which so far has made little use of notions like event and frame. The incentive for work in this tradition derives from sociologists' attempts to find alternatives to the symbolic interactionists' measures of small group interaction, which relied on statistical counts of *a priori* content categories. Such categories had repeatedly been criticized as having no demonstrable relationship to actual behavior. In a brilliant series of experiments, Garfinkel (1967, 1972) demonstrates that social knowledge cannot be adequately characterized in the form of statistically countable, abstract categories such as scalar ratings of role, status or personality characteristics. He argues that social knowledge is revealed in the process of interaction itself and that interactants create their own social world by the way in which they behave. He then goes on to suggest that sociology should concentrate on describing the mechanisms by which this is done in what he calls "naturally organized activities," rather than in staged experiments or interview elicitations.

Sacks and his collaborators (Garfinkel & Sacks 1970, Sacks 1972, Schegloff 1972, Sacks, Schegloff & Jefferson 1974, Turner 1974) were the first systematically to focus on conversation as the simplest instance of a naturally organized activity, and attempt to study the process of conversational management *per se* without making any *a priori* assumptions about social and cultural background of participants. Their research concentrated on isolating strategies of effecting speaker change, opening and closing conversations, establishing

semantic relations between utterances, signalling asides and sequences, and otherwise controlling and channeling the course of an interaction.

The picture of everyday conversation that emerges from this work is one of a dynamic interactive flow marked by constant transitions from one mode of speaking to another: shifts from informal chat to serious discussion, from argument to humor, or narrative to rapid repartee, etc. In other words, speech routines, which when seen in speech act terms constitute independent wholes, here serve as discourse strategies integrated into and interpreted as part of the broader task of conversational management. Conversational analysis over the last few years has demonstrated beyond question that not only formally distinct speech events but all kinds of casual talk are rule governed. It is through talking that one establishes the conditions that make an intended interpretation possible. Thus to end a conversation, one must prepare the ground for an ending; otherwise, the ending is likely to be misunderstood. Or to interpret an answer, one must be able to identify the question to which that answer is related. To understand a pun, one must be able to retrieve, re-examine and reinterpret sequences that occurred earlier in an interaction. Sequentiality, i.e. the order in which information is introduced and the positioning or locating of a message in the stream of talk, is clearly of great importance in interpreting daily conversation. The mechanisms which underlie speaker-listener coordination can be studied empirically by examining recurrent strategies, the responses they elicit, and the ways in which they are modified as a result of those responses.

One of Sacks' key contributions to conversational analysis is his recognition that principles of conversational inference are quite different from rules of grammar. Rather than 'rule,' he uses the term 'maxim,' which is reminiscent of Grice's (1975) notion of implicature, to suggest that interpretations take the form of preferences rather than obligatory rules. The point is that at the level of conversation, there are always many possible alternative interpretations, many more than exist at the level of sentence grammar. Choice among these is constrained by what the speaker intends to achieve in a particular interaction, as well as by expectations about the other's reactions and assumptions. Yet once a particular interpretation has been chosen and accepted it must be followed. That is, an interpre-

tive strategy holds until something occurs in the conversation to make participants aware that a change in strategy is indicated. Interpretations are thus negotiated, repaired and altered through interactive processes rather than unilaterally conveyed.

Conversational analysts were the first to provide systematic evidence for the cooperative nature of conversational processes and to give interactional substance to the claim that – to use Halliday's expression – words have both relational and ideational significance. The perspective they have developed is therefore crucial to the study of verbal encounters. Yet their work does not account for the linguistic bases of conversational cooperation. Theoretical writings in this tradition see the post-Chomskian concern with grammatical rules as merely another instance of the normative sociological paradigm they have been trying to overcome. When linguists' findings are discussed it is mainly to point out their limitations (Cicourel 1974). Yet in much of the empirical work of conversational analysts referential meanings that assume sharing of contextualization strategies are taken for granted.

This view of language has serious limitations which affect both the validity of the analysts' attempts to capture participants' interpretive processes and the social import of their work. In order to account for inter-speaker differences in background knowledge, a sociolinguist needs to know how speakers use verbal skills to create contextual conditions that reflect particular culturally realistic scenes. How are speakers' grammatical and phonological abilities employed in this? For example, if regular speaker change is to take place, participants must be able to scan phrases to predict when an utterance is about to end. They must be able to distinguish between rhetorical pauses and turn relinquishing pauses. Although overlap is an integral part of interaction, conversational cooperation requires that interactional synchrony be maintained so that speakers cannot be interrupted at random. To follow the thematic progression of an argument, moreover, and to make one's contribution relevant, one must be able to recognize culturally possible lines of reasoning. It is therefore necessary to show how strategies of conversational management are integrated into other aspects of speakers' linguistic knowledge.

Recovering background knowledge

To this end, in what follows several examples of actual conversation

will be examined to illustrate the limitations of the three traditions discussed – ethnography of communication, discourse and conversational analysis – and to suggest a way of utilizing the insights provided by these three traditions to build a more comprehensive theory of conversational inference. These examples are representative of a much larger body of data we have collected, both by chance, as in these examples, and in connection with systematic programs. The first examples reflect exchanges which any native speaker of English would be able to interpret. The fourth constitutes an inter-ethnic encounter, and shows some of the inferential processes that underlie misinterpretation of intent.

- (1) The first incident was recorded while I was sitting in an aisle seat on an airplane bound for Miami, Florida. I noticed two middle aged women walking towards the rear of the plane. Suddenly I heard from behind, "Tickets, please! Tickets, please!" At first I was startled and began to wonder why someone would be asking for tickets so long after the start of the flight. Then one of the women smiled toward the other and said, "*I told* you to leave him at home." I looked up and saw a man passing the two women, saying, "*Step* to the rear of the bus, please."

Americans will have no difficulty identifying this interchange as a joke, and hypothesizing that the three individuals concerned were probably travelling together and were perhaps tourists setting off on a pleasure trip. What we want to investigate is what linguistic and other knowledge forms the basis for such inferences, and to what extent this knowledge is culturally specific.

The initial utterance, "Tickets, please," was repeated without pause and was spoken in higher than normal pitch, with more than usual loudness, and in staccato rhythm. For this reason it sounded like an announcement, or like a formulaic phrase associated with travel situations. My first inkling that what I heard was a joke came with the woman's statement to her friend, "*I told* you to leave him at home." Although I had no way of knowing if the participants were looking at each other, the fact that the woman's statement was perfectly timed to follow the man's utterance was a cue that she was responding to him, even though her comment was addressed to a third party. Furthermore, the stress on "told" functioned to make her statement sound like a formulaic utterance, contributing to the hypothesis that she and he were engaging in a similar activity. If either the man or the woman had uttered their statements in normal

pitch and conversational intonation, the connection between them might not have been clear. Only after I was able to hypothesize that the participants were joking, could I interpret their utterances. My hypothesis was then confirmed by the man's next statement, "Step to the rear of the bus, please." This was also uttered in announcement style. In retrospect, we may note that both of the man's utterances were formulaic in nature, and thus culturally specific and context bound. He was exploiting the association between walking down an aisle in a plane and the similar walk performed by a conductor on a train or a bus. In identifying the interaction as a joke, I was drawing on the same situational knowledge, as well as on my awareness of the fact that tourists bound for Miami are likely to engage in such joking.

Suprasegmental and other surface features of speech are often crucial to identifying what an interaction is about. When seen in isolation, sentences can have many intonation and paralinguistic contours, without change in referential meaning. As was pointed out in previous chapters, the prevalent view is that these suprasegmental features add expressive overtones to basic meanings conveyed by core linguistic processes, i.e. the signs by which listeners recognize these overtones tend to be seen as language-independent. The incident provides evidence for our claim that prosody is essential to conversational inference. The identification of specific conversational exchanges as representative of socio-culturally familiar activities is the process I have called 'contextualization' (chapter 6). It is the process by which we evaluate message meaning and sequencing patterns in relation to aspects of the surface structure of the message, called 'contextualization cues.' The linguistic basis for this matching procedure resides in 'co-occurrence expectations,' which are learned in the course of previous interactive experience and form part of our habitual and instinctive linguistic knowledge. Co-occurrence expectations enable us to associate styles of speaking with contextual presuppositions. We regularly rely upon these matching processes in everyday conversation. Although they are rarely talked about and tend to be noticed only when things go wrong, without them we would be unable to relate what we hear to previous experience.

(2) This incident was recorded at the end of a helicopter flight from a Bay Area suburb to San Francisco airport. The cabin attendant whose seat was squeezed in among the half dozen passengers all grouped together

in the center of the aircraft picked up the microphone and addressed the group:

We have now landed at San Francisco Airport. The local time is 10.35. We would like to thank you for flying SFO Airlines, and we wish you a happy trip. Isn't it quiet around here? Not a thing moving.

Here prosody and rhythm serve to distinguish two quite separate activities. The last two sentences were preceded by a slight pause and marked by lowering of pitch, increase in tempo and more pronounced intonational contouring. The passengers identified it as a personal remark which, although spoken through the microphone, was not part of the announcement. But simply to note that the attendant has engaged in two distinct speech activities does not explain the interactive facts. An announcement is a unilateral statement, which, particularly in a suburban flight, does not require listener response. It is understood that it is being made to conform to the legal requirements and does not reflect the speaker's opinion. In a personal statement, however, speakers assume responsibility for their words and may expect a response. In the present case several passengers reacted by nodding. One person asked why it was so quiet whereupon the attendant replied that cargo personnel were on strike. The incident illustrates the hierarchical nature of inferential processes, in which higher level assessments feed into our interpretation of component utterances and affect listener responses.

Signalling of frames by a single speaker is not enough. All participants must be able to fit individual contributions into some overall theme roughly corresponding to a culturally identifiable activity, or a combination of these, and agree on relevant behavioral norms. They must recognize and explicitly or implicitly conform to others' expectations and show that they can participate in shifts in focus by building on others' signals in making their own contributions.

One common way in which conversational cooperation is communicated and monitored by participants is through what Yngve (1970) calls "back channel signals": interjections such as, "O.K.," "right," "aha," or nods or other body movements. Other signs of cooperation are implied indirectly in the way speakers formulate responses, i.e. in whether they follow shifts in style, agree in distinguishing new from old or primary from secondary information, or in

judging the quality of interpersonal relationships implied in a message, and know how to fill in what is implied but left unsaid or what to emphasize or de-emphasize.

(3) This is another striking example of how contextualization works and enters into interpretation of intent. The incident was observed at a luncheon counter, where the waitress behind the counter was talking with a friend seated at the counter:

Friend: I called Joe last night.
 Waitress: You did? Well what'd he say?
 Friend: Well, hi!
 Waitress: Oh yeah? What else did he say?
 Friend: Well he asked me out of course.
 Waitress: Far out!

To participate in this exchange, the waitress, apart from having to rely on socio-cultural schemata about dating situations, must recognize that the first statement, which seems complete on the surface, is actually the lead-in for a story that she is expected to help elicit. Further, she must know that "called" refers to a telephone call; she must know who Joe is; and she must realize that the call was not routine but had special meaning for her friend. Her reply "You did?" with exaggerated intonation contour and vowel elongation on "did," implicitly acknowledges all this. She then demonstrates that she has an idea of what's coming next in the story by her prompt, "Well what'd he say?"

Note that the friend's response gives the main point of her story, but the meaning is almost entirely conveyed not by the content of what is said but by *how* it is said. This is communicated largely through prosody. In other words, participants must infer that the fall rise intonation on greetings such as "Hi" may signal surprise mixed with pleasure. Such intonation contours become meaningful through recurrent association with certain speech activities. Only if we know this, and are acquainted with the relevant conventions, can we interpret the speaker's use of "of course" in her subsequent comment.

How can empirical examinations of inferencing in examples such as these be used in developing a more general theory of what accounts for both shared and culturally specific aspects of interpretive processes? It seems clear that each of the three traditions we have discussed has something of importance to contribute. At the level of

ethnographic description, verbal behavior in all societies can be categorized in terms of speech events: units of verbal behavior bounded in time and space. Events vary in the degree to which they are isolable. They range from ritual situations where behavior is largely predetermined to casual everyday talk. Yet all verbal behavior is governed by social norms specifying participant roles, rights and duties vis-à-vis each other, permissible topics, appropriate ways of speaking and ways of introducing information. Such norms are context and network specific, so that the psycholinguistic notion of individuals relying on their own personal knowledge of the world to make sense of talk in context is an oversimplification which does not account for the very real interactive constraints that govern everyday verbal behavior.

When events are named, such names are regularly employed in members' narrative reports in sentences such as "We attended a lecture," "They were making a joke." Events also serve as labels for the constellations of norms by which verbal behavior is evaluated, so that someone commenting on the helicopter announcement might say "They said it as part of a formal announcement and didn't mean it personally."

But no one could argue that the descriptions of time bound event sequences can account for the interpretive issues discussed here. Apart from the fact that verbal interchanges rarely take the form of set, isolable routines and that event labels often do not characterize what is actually intended, there is the problem of inducing potential conversationalists to participate. Conversational cooperation, as we have argued following Grice, is always cooperation for some purpose, which means that participants must have at least some idea of the likely outcomes before they commit themselves to an interaction. Where potential outcomes are not agreed upon in advance they must be negotiated through talk. Information about interactive goals, therefore, has to be conveyed before enough has transpired to make a sequential description possible. Example (3), for instance, could in retrospect be described as a personal narrative, but the listener might not have listened and given the responses she did give had she not predicted that narrating was intended. Some abstract cognitive concept like the discourse analyst's schema is therefore called for. But schemata, as our data tell us, cannot simply refer to knowledge of the physical world. In fact I would argue that a cognitive approach to

discourse must build on interaction. It must account for the fact that what is relevant background knowledge changes as the interaction progresses, that interpretations are multiply embedded and that, as Goffman (1974) has shown, several quite different interactions are often carried on at the same time. We need a semantic concept closer perhaps to Frake's (1972) and Agar's (1975) use of the term 'event' defined in terms of communicative goals. For this purpose, we will use the term 'speech activities' (Levinson 1978).

A speech activity is a set of social relationships enacted about a set of schemata in relation to some communicative goal. Speech activities can be characterized through descriptive phrases such as "discussing politics," "chatting about the weather," "telling a story to someone," and "lecturing about linguistics." Such descriptions imply certain expectations about thematic progression, turn taking rules, form, and outcome of the interaction, as well as constraints on content. In the activity of discussing, we look for semantic relationships between subsequent utterances, and topic change is constrained. In the activity of chatting, topics change freely, and no such expectations hold. Lecturing, in turn, implies clear role separation between speaker and audience and strong limitations on who can talk and what questions can be asked.

Note that the descriptive phrases we use for speech activities contain both a verb and a noun which suggests constraints on content. Verbs alone, or single nouns such as "discussion," or "lecture," are not sufficient. Activities are not bounded and labelable entities but rather function as guidelines for the interpretation of events which show certain general similarities when considered in the abstract but vary in detail from instance to instance. One should not expect to be able to find a limited set of speech activities.

Although speech activities are thus not precisely listable, they are the means through which social knowledge is stored in the form of constraints on action and on possible interpretation. In verbal interaction social knowledge is retrieved through co-occurrence expectations of the type we have discussed. Distinctions among such activities as chatting, discussing and lecturing exist in all cultures, but each culture has its own constraints not only on content but also on the ways in which particular activities are carried out and signalled. Even within a culture, what one person would identify as "lec-

turing," another might interpret as "chatting with one's child," and so on. What the usual labels reflect are Wittgensteinian family resemblances rather than analytical categories.

Since speech activities are realized in action and since their identification is a function of ethnic and communicative background special problems arise in a modern society where people have widely varying communicative and cultural backgrounds. How can we be certain that our interpretation of what activity is being signalled is the same as the activity that the interlocutor has in mind, if our communicative backgrounds are not identical? It is here that the work on conversational synchrony discussed in chapter 6 takes on special importance.

In the spirit of this work, I would like to suggest that the signalling of speech activities is not a matter of unilateral action but rather of speaker-listener coordination involving rhythmic interchange of both verbal and nonverbal signs. In other words, a successful interaction begins with each speaker talking in a certain mode, using certain contextualization cues. Participants, then, by the verbal style in which they respond and the listenership cues they produce, implicitly signal their agreement or disagreement; thus they 'tune into' the other's way of speaking. Once this has been done, and once a conversational rhythm has been established, both participants can reasonably assume that they have successfully negotiated a frame of interpretation, i.e. they have agreed on what activity is being enacted and how it is to be conducted. At this point, a principle of strategic consistency takes over similar to that which Sacks (1972) refers to as the 'parsimony principle.' Speakers continue in the same mode, assigning negotiated meanings to contextualization cues, until there is a perceptible break in rhythm, a shift of content and cues, or until a mismatch between content and cues suggests that something has gone wrong.

It is clear, looking at conversation in this way, that if conversational inference is a function of identification of speech activities, and if speech activities are signalled by culturally specific linguistic signs, then the ability to maintain, control and evaluate conversation is a function of communicative and ethnic background.

The next example illustrates some of the inferential problems that arise when different background expectations are employed in the interpretation of a single message.

(4) The incident took place in London, England, on a bus driven by a West Indian driver/conductor. The bus was standing at a stop, and passengers were filing in. The driver announced, "Exact change, please," as London bus drivers often do. When passengers who had been standing close by either did not have money ready or tried to give him a large bill, the driver repeated, "Exact change, please." The second time around, he said "please" with extra loudness, high pitch, and falling intonation, and he seemed to pause before "please." One passenger so addressed, as well as others following him, walked down the bus aisle exchanging angry looks and obviously annoyed, muttering, "Why do these people have to be so rude and threatening about it?"

Was the bus driver really annoyed? Did he intend to be rude, or is the passengers' interpretation a case of cross-cultural misunderstanding? The cues in the example given here are largely prosodic. I will attempt to show how prosody and paralinguistic cues function in signalling frames of interpretation.

We can assume that English speaking listeners rely upon their native presuppositions to segment the passage into relevant processing units and to retrieve information not overtly expressed through lexical means. According to this system the utterance in question could be spoken as a single tone group:

(5) Exact change please //

as it was the first time the driver said it, or as two tone groups:

(6) Exact change / please //

as he said it the second time. To treat "please" as a distinct information unit implies that it is to be given special attention or emphasis and this is a possible option. Tone grouping by itself therefore is not an issue here. However accent placement and tune do create problems. One might argue that in a short, syntactically simple utterance such as the present one, the accent would ordinarily fall on "change." But even in simple sentences accent placement is affected by activity-specific expectations. If I say:

(7) I'm giving my paper //

"paper" is accented because it reflects the expected point of information focus. However in:

(8) I'm cancelling my paper //

the verb is normally accented since "cancelling" is not considered a customary activity in relation to paper giving.

In the bus driver case, requesting exact change is customary so that the accent on "change" would be expected. But note that the politeness tag "please" is also accented and carries a falling tone. This goes counter to English prosodic conventions which associate falling tones with definiteness and finality, while rising tones, among other things, count as tentative and therefore tend to sound more polite. The interpretive effect here is the reverse from what happens when phrases like "This is nice" are given a rising tone to convey that a previous statement or pre-existing attitude is being questioned. "Please" spoken with a falling tone by contrast implies annoyance at something the listener did or is likely to do.

Consider now the driver's second utterance, where "change" with falling tone is followed by "please" marked by a separate tone group and by extra loudness and a shift to a higher fall. A speaker of British English in repeating this utterance could optionally (a) place the accent on "change" or (b) split the sentence into two tone groups, as the driver did. In (a) the normal interpretation would be "I said, change." In (b) setting off "please" would highlight the directness of the request. Directness in public situations is likely to cause offense so that the mitigating effect of a rising or falling rising tune becomes even more important. Since the driver here seems to be doing just the opposite, the interpretation of rudeness is natural for listeners who rely on English contextualization conventions to infer motivation.

Yet, in order to determine whether the conclusion that the driver was being rude corresponds to West Indian contextualization conventions, we need to look at how prosodic and paralinguistic cues normally function in West Indian conversation. Examination of the contextualization practices employed in our recordings of West Indian Londoners conversing in informal in-group settings, suggests that their use of prosody and paralinguistics is significantly different from that of British English or American English speakers. For example, syntactic constraints on the placement of tone group boundaries differ. West Indians can split a sentence into much smaller tone group units than British English speakers can. In addition, their use of rising tune to indicate the contrast between tentativeness and definiteness and inter-clausal cohesion is much more restricted. Moreover, once a tone group boundary has been established, nuc-

leus placement within such a tone group must be on the last content word of that tone group regardless of meaning. In contrast to other forms of English, nucleus placement is syntactically rather than semantically constrained. The bus driver's accent on "please" can therefore be seen as an automatic consequence of tone grouping, not a matter of conscious choice. Finally, pitch and loudness differences do not necessarily carry expressive connotations. They are regularly used to indicate emphasis without any overtones of excitement or other emotion. To illustrate, in the course of an ordinary, calm discussion, one speaker said:

(9) He was selected / *mainly* / because he had a degree //

The word "mainly" was separated by the tone group boundaries and set off from the rest of the sentence by increased pitch and loudness. The overall context within which that sentence occurs shows that the word "mainly" was used contrastively within a line of reasoning which argued that having practical experience was as important as formal education. Our conclusion is that the West Indian bus driver's "Exact change / please //" was simply his accustomed way of emphasizing the word "please," corresponding to the British option (b) above. Therefore, his intention was, if anything, to be polite.

To summarize then, we conclude that the conversational inference processes we have discussed involve several elements. On the one hand is the perception of contextualization cues. On the other is the problem of relating them to other signalling channels. Interpretation, in turn, requires first of all judgements of expectedness and then a search for an interpretation that makes sense in terms of what we know from past experience and what we have perceived. We can never be certain of the ultimate meaning of any message, but by looking at systematic patterns in the relationship of perception of surface cues to interpretation, we can gather strong evidence for the social basis of contextualization conventions and for the signalling of communicative goals.

The linguistic character of contextualization cues is such that they are uninterpretable apart from concrete situations. In contrast to words or segmental morphemes which, although ultimately also context-bound, can at least be discussed in isolation, listed in dictionaries and explained in grammars, contextualization phenomena are impossible to describe in abstract terms. The same sign may

indicate normal information flow under some conditions and carry contrastive or expressive meanings under others. We are faced with a paradox. To decide on an interpretation, participants must first make a preliminary interpretation. That is, they listen to speech, form a hypothesis about what routine is being enacted, and then rely on social background knowledge and on co-occurrence expectations to evaluate what is intended and what attitudes are conveyed.

What distinguishes successful from unsuccessful interpretations are not absolute, context-free criteria of truth value or appropriateness, but rather what happens in the interactive exchange itself, i.e. the extent to which proffered context bound inferences are shared, reinforced, modified or rejected in the course of an encounter. Ultimately, of course, anything that is said is subject to being evaluated in terms of social norms and established criteria of truthfulness and rationality. But the contextual criteria in terms of which these judgements are made are often quite different from those applying to conversational inference and this has important implications for our understanding of culture and communication.